journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Information quality measurement of medical encoding support based on usability

John Puentes[a,*], Julien Montagner[a], Laurent Lecornu[a], Jean-Michel Cauvin[b,c]

[a] Institut Mines-Telecom; Telecom Bretagne, UEB; Dpt Image et Traitement de l'Information, Brest, France
[b] Département d'Information Médicale CHRU, Brest, France
[c] Inserm UMR 1101 LaTIM, Brest, France

ARTICLE INFO

Article history:
Received 24 January 2013
Received in revised form
18 July 2013
Accepted 20 July 2013

Keywords:
Information quality assessment
Medical information analysis
Clinical encoding
Usability
Decision support
Hospital information systems

ABSTRACT

Medical encoding support systems for diagnoses and medical procedures are an emerging technology that begins to play a key role in billing, reimbursement, and health policies decisions. A significant problem to exploit these systems is how to measure the appropriateness of any automatically generated list of codes, in terms of fitness for use, i.e. their quality. Until now, only information retrieval performance measurements have been applied to estimate the accuracy of codes lists as quality indicator. Such measurements do not give the value of codes lists for practical medical encoding, and cannot be used to globally compare the quality of multiple codes lists. This paper defines and validates a new encoding information quality measure that addresses the problem of measuring medical codes lists quality. It is based on a usability study of how expert coders and physicians apply computer-assisted medical encoding. The proposed measure, named ADN, evaluates codes Accuracy, Dispersion and Noise, and is adapted to the variable length and content of generated codes lists, coping with limitations of previous measures. According to the ADN measure, the information quality of a codes list is fully represented by a single point, within a suitably constrained feature space. Using one scheme, our approach is reliable to measure and compare the information quality of hundreds of codes lists, showing their practical value for medical encoding. Its pertinence is demonstrated by simulation and application to real data corresponding to 502 inpatient stays in four clinic departments. Results are compared to the consensus of three expert coders who also coded this anonymized database of discharge summaries, and to five information retrieval measures. Information quality assessment applying the ADN measure showed the degree of encoding-support system variability from one clinic department to another, providing a global evaluation of quality measurement trends.

1. Introduction

The fundamental goal of medical encoding is to identify diagnosis related groups of patients and determine the corresponding healthcare expenses, billing, and reimbursement. Medical encoding is used in addition to record diseases morbidity and causes of mortality. This encoded information has become increasingly important, given its impact on medical activities evaluation at various levels of health

* *Corresponding author*. Tel.: +33 2290013 39.
E-mail address: John.Puentes@telecom-bretagne.eu (J. Puentes).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.07.018

organizations. Moreover, encoding relevance affects patient management, along with epidemiologic, safety, research, and health policies decisions [1]. Medical codes are assigned to define diagnoses and procedures of each care episode that occurred during an inpatient stay. Codes represent [2,3]: main and secondary diagnoses, complications, comorbidities, primary and secondary procedures. Currently, most of medical encoding is carried out in two distinct manual manners:

- By expert coders who, without having any particular ancillary knowledge of the specific patient history, produce lists of codes that are considered to be exhaustive.
- By physicians who code essential aspects of the care episodes having some knowledge of the specific patient history, but usually generate a subset of the codes list produced by expert coders, because of practical restrictions (mainly focus on current diseases, limited awareness of encoding guidelines, and short available time).

In both cases medical encoding is expensive, taking much more time for the experts than for the physicians. The main reason is that, in addition to the patient record human coders have to examine hundreds of candidate codes in encoding references, to define the codes list that represents a given inpatient stay. Nevertheless, the pertinence of resulting codes sets strongly depends on the variable coders' expertise, which often produces inacceptable results like under or over encoding [1,4].

An emerging alternative is computer-assisted medical encoding technology. It analyses available patient information to automatically generate a list of most pertinent medical codes. Thereafter, coders select the appropriate codes corresponding to a specific inpatient stay. In more than a decade, several approaches have been developed to produce the corresponding encoding support lists, using varied types of input information. These studies considered, among others: extraction of semantic labels from documents [5,6]; matching of structured encoding forms and parsed clinical information [7]; correlation with precedent encoding results [8]; use of encoding rules [9]; codes linked to specific keywords [10]; a combination of an encoding classification with ontologies and natural language [11].

A significant problem with this kind of technology is how to measure the proposed codes lists appropriateness in terms of fitness for use, i.e. quality, according to: the distribution of correct and incorrect codes along the list, the amount of expected correct codes, the observation windows, and the variable list length. That information quality measurement should assert the practical value of any codes list, in a suitable manner adapted to the different encoding practices of hospitals and countries [12–15]. Automatically generated codes lists represent nevertheless information of variable quality, depending on the quality of input data. Such quality is complex to determinate particularly on the heterogeneous data sets of any hospital information system (HIS). Otherwise, the pertinence of codes lists produced by an encoding support system is conventionally estimated by comparing suggested codes, with a reference encoding done by an expert coder. Nevertheless, this approach does not provide any clues about the lists' value for medical encoding. Information retrieval performance measurements have been also used to estimate accuracy as quality indicator. These measurements do not give either the value of codes lists in the sense of their adequacy to encoding practices, and cannot be used to globally compare the quality of multiple codes lists. This paper thus addresses the problem of how to measure the appropriateness of an automatically generated codes list, in terms of fitness for use. We define and validate a new information quality measure that copes with limitations of previously applied measures. It is based on how expert coders and physicians make use of computer-assisted medical encoding. The proposed measure, named ADN, evaluates codes Accuracy, Dispersion, and Noise in the whole generated list, independently of its content and length. According to the ADN measure, the information quality of a codes list is fully represented by a single point within a normalized triangular space, partitioned by iso-quality lines. Moreover, our approach is reliable to examine and compare, using a unique scheme, the information quality of hundreds of codes lists, showing their practical value for encoding.

1.1. *Background*

In general, data quality analyses have been focused on well-known issues (wrong, missing or unusable data) produced by both humans and systems, at any stage of the data existence cycle. These issues accumulate generating varied complex functional problems [16,17]. Even if multiple approaches have been proposed to systematically identify, characterize, and correct inconsistencies produced by deficient data [18–22], information quality assessment remains a central and unexplored challenge [23–26]. Furthermore, encoding information quality measurement is particularly necessary when a regular audit of the associated HIS applications cannot be done, due to functional constraints.

Examined documents to produce medical codes vary from the whole patient record to discharge summaries. Included data and information are expected to be accurate, i.e. truly represent the element each value was intended for. As a consequence, only accuracy has been commonly considered as analogous to quality in the medical domain. For more than 30 years precision (*Pr*) and recall (*Rc*) have been the main applied information retrieval measures [27,28], along with complementary related evaluations [29,30]. For any retrieval system, its overall performance is comparatively determined using a set of precision-recall curves [31]. Rank measures of relevant documents can be calculated considering average weighed precision, by means of R-Measure and Q-measure [32,33]. Additionally, to handle incomplete information, binary relevance judgment defines globally a preference relation with respect to relevant documents [34]. Until now, only measurements used for information retrieval performance evaluation have been applied to estimate medical data and information quality. Based on some of these approaches, accuracy of computer-based patient records data was estimated applying two complementary measures [35]: *Cr* – correctness (proportion of correct data) and *Cm* – completeness (proportion of rightly recorded data), defined as:

$$Cr = \frac{tp}{tp + fp} = Pr = PPV \tag{1}$$